Multiple Random Forests Based Intelligent Location of Single-phase Grounding Fault in Power Lines of DFIG-based Wind Farm

Yongli Zhu and Hua Peng

***Abstract*—To address the problems of wind power abandonment and the stoppage of electricity transmission caused by a short circuit in a power line of a doubly-fed induction generator (DFIG) based wind farm, this paper proposes an intelligent location method for a single-phase grounding fault based on a multiple random forests (multi-RF) algorithm. First, the simulation model is built, and the fundamental amplitudes of the zero-sequence currents are extracted by a fast Fourier transform (FFT) to construct the feature set. Then, the random forest classification algorithm is applied to establish the fault section locator. The model is resampled on the basis of the bootstrap method to generate multiple sample subsets, which are used to establish multiple classification and regression tree (CART) classifiers. The CART classifiers use the mean decrease in the node impurity as the feature importance, which is used to mine the relationship between features and fault sections. Subsequently, a fault section is identified by voting on the test results for each classifier. Finally, a multi-RF regression fault locator is built to output the predicted fault distance. Experimental results with PSCAD/EMTDC software show that the proposed method can overcome the shortcomings of a single RF and has the advantage of locating a short hybrid overhead/cable line with multiple branches. Compared with support vector machines (SVMs) and previously reported methods, the proposed method can meet the location accuracy and efficiency requirements of a DFIG-based wind farm better.**

***Index Terms*—Doubly-fed induction generator (DFIG) based wind farm, power line, multiple random forests (multi-RF), single-phase grounding fault, fault location.**

I. Introduction

THE proportion of wind power generation in power systems is gradually increasing during the low carbon and intelligent energy transition period to achieve carbon neutrality. As most of the power collection lines in wind farms are located in mountainous areas or seas with harsh environmental conditions, they are prone to suffer from lightning strikes and the windage of insulators, resulting in short-circuit faults. Among these faults, single-phase grounding faults account for approximately 80% in doubly-fed induction generator (DFIG) based wind farms, with the highest probability of occurrence. A power line in a wind farm is usually composed of many wind turbine branches and short-distance overhead and cable lines, which increase the difficulty of locating a fault [1]. Therefore, the rapid and accurate location of faults in power collection lines can effectively guide maintenance personnel to repair them. Moreover, the interruption time in electricity transmission in these lines is decreased, and the power generation loss of the wind turbines of faulty lines can be significantly reduced, which increases the economic benefits of the wind farm.

The electric system of a wind farm is a typical multi-terminal power generation and transmission system. At present, there are a few studies on fault location for this type of network, resulting in a lack of effective diagnostic techniques and means of location. Fault location approaches can be mainly categorized into traveling wave approaches [2]-[7], fault analysis approaches [8]-[12], and artificial intelligence approaches [13]-[18]. Traveling wave approaches are simple, but there are multi-branch lines and hybrid lines in the distribution network and wind farm, which make it difficult to recognize the traveling wave head and determine the wave speed. Because the power lines in wind farms are short and the sampling frequency of the traveling wave is restricted, the fault location error increases. Considering the difference in the traveling wave velocity in overhead and cable sections, the location of a fault in a hybrid distribution system is determined on the basis of the time difference in the initial traveling wave [6]. In [7], a traveling wave location method is presented using the discrete wavelet transform (DWT) for a hybrid multi-terminal transmission system that interconnects an offshore wind farm and the main grid. However, its location performance is reduced to a certain degree when a fault occurs near the joint points and ending terminals [7].

Fault analysis methods have a high stability. Nevertheless, the location precision is sensitive to variations in the fault resistances and system parameters. In [11], an active signal injection scheme is designed based on a soft open point. The low-frequency amplitudes of the zero-sequence currents extracted by a fast Fourier transform (FFT) are compared with the value of a current threshold in turn to realize the accu-

Manuscript received: August 25, 2021; revised: November 23, 2021; accepted: January 25, 2022. Date of CrossCheck: January 25, 2022. Date of online publication: March 2, 2022.

This work was supported in part by the National Natural Science Foundation of China (No. 51677072).

Y. Zhu and H. Peng (corresponding author) are with the School of Electrical and Electronic Engineering, North China Electric Power University, Baoding 071003, China (e-mail: yonglipw@163.com; penghua890415@126.com).

DOI: 10.35833/MPCE.2021.000590

rate judgment of a fault section in a resonant grounding distribution network. As explained in [11], the scheme needs to search the fault section of feeders step by step. Consequently, it has significant negative impacts such as the high redundancy during computation and the low location efficiency. The scheme in [12] is suitable for dual-end and T-connected transmission lines, but it is difficult to adapt to fault location for the wind farm power lines of hybrid short conductors and multiple branches, which cause a large number of calculations with multiple iterations.

The construction of a digitally driven smart wind farm has opened an era of artificial intelligence in the field of wind power. Hence, it has become a trend that intelligent algorithms are selected for fault location for wind farms with multiple branches and hybrid short lines. In [17], the norm of the detail coefficients, which are extracted by a DWT from the voltage signal measured at the midpoint of the DC/DC converter, is utilized as a feature vector. Subsequently, a pattern recognition method based on a three-layer feed-forward artificial neural network is carried out to identify the correct fault location in a photovoltaic system. In [18], the AC current and internal circulation current are analyzed when an open-circuit fault of a submodule occurs in different bridge arms in a modular multi-level converter (MMC). Then, a mixed-kernel support tensor machine (STM) is applied to locate the faulty MMC, which achieves a better accuracy than that of a single-kernel STM [18].

The random forest (RF) algorithm [19]-[22] is an excellent ensemble learning algorithm based on the bagging and random subspace methods, which takes into account both the prediction accuracy and efficiency. Furthermore, the RF algorithm has been widely used in the fields of power quality (PQ) complex disturbance recognition [20], multi-source partial discharge (PD) diagnosis of a transformer [21], and prediction of daily enterprise electricity consumption [22]. In [20], PQ disturbances are broadly classified into two classes on the basis of the FFT and the calculation of the number of peaks above a threshold. A modified S-transform is then used for feature extraction, and at the same time, the RF algorithm is used to effectively assess the PQ. Even at a noise level 20 dB, its classification result is still accurate, with a result of 99.61%. In [21], the S-transform aided RF algorithm is presented to detect the location of the PD, which employs the signature of an optical sensor, and the accuracy rate obtained for recognizing the location of the PD is 95.6%. In [22], time-frequency features are extracted by an ensemble empirical mode decomposition FFT (EEMD-FFT). On this basis, the RF algorithm is used to predict the daily enterprise electricity consumption, which outperforms other predictors, e. g., the back propagation neural network (BPNN) and the least-squares support vector machine (LSSVM). Accordingly, the RF algorithm has the following significant advantages: ① high generalization ability and noise immunity; ② capability of preventing overfitting; and ③ applicability to the rapid processing of large-scale power data.

With the background of wind power big data applications and smart wind farm construction, an intelligent single-phase grounding fault location method for power lines based on multiple random forests (multi-RF) is proposed in this paper. A 110 kV/35 kV simulation model is built, and the fundamental amplitudes of the zero-sequence currents are extracted by the FFT to construct the feature set. Furthermore, an RF fault section locator is introduced to calculate the mean decrease in the value of the node impurity, which is used to select features for quantitative analysis and sorting. A fault section is accurately located through the voting mechanism of the RF fault section locator. On this basis, a multi-RF regression fault locator is established to output the predicted fault distance. The experimental results obtained using PSCAD/EMTDC software show that the proposed method is not affected by the fault location, transition resistance, or noise, and can be applied to locate faults in DFIG-based wind farms.

A few methods have been used for power line fault location in DFIG-based wind farms. Compared with these methods, the proposed method has the following characteristics. First, it can cope with the fault locations of shorter hybrid power lines. Second, the dual-ended traveling wave method cannot adapt to the structure of a power network with wind farms, whose lines have multiple generator branches. Compared with this method, the proposed method adopts the fundamental amplitudes of the zero-sequence currents of a power line as features; hence, there is no location error caused by the clock synchronization problem. In addition, a high sampling frequency is not required; thus, the running speed and processing efficiency are naturally good. Third, compared with the impedance method, the proposed method avoids the need to solve complex equations, which makes it possible to quickly and accurately locate fault points without the problems of pseudo roots and pseudo fault points of the impedance method.

The remainder of this paper is organized as follows. Section II describes the topological structure of the electric system of a large-scale wind farm. Section III presents the simulation model of the DFIG-based wind farm and sets the conditions for generating the training set. Section IV presents the methodology for fault characteristic analysis and fault feature extraction. Section V proposes an approach for constructing a multi-RF fault locator for DFIG-based wind farms. Section VI introduces the design of multi-RF fault locator for power lines of DFIG-based wind farms. Section VII analyzes and discusses the experimental results obtained using the proposed method. Finally, Section VIII concludes the paper.

II. Topological Structure of Electric Power System of a Large-scale Wind Farm

The topological structure of the electric power system of a large-scale wind farm, which is composed of wind turbines, power lines, and a booster substation, is shown in Fig. 1. At present, the system has the following general structural characteristics.

1) From the overall perspective, it is a typical radiation network, which adopts the connection form of a chain.

2) Its main voltage level is 35 kV. In order to ensure its safety and reliability, a zig-zag grounding transformer is usually utilized to constitute its artificial neutral point.

3) Each associated power line is composed of hybrid overhead/cable sections and is connected to multiple wind turbines. Moreover, the lengths between adjacent wind turbines

are different and usually very short.

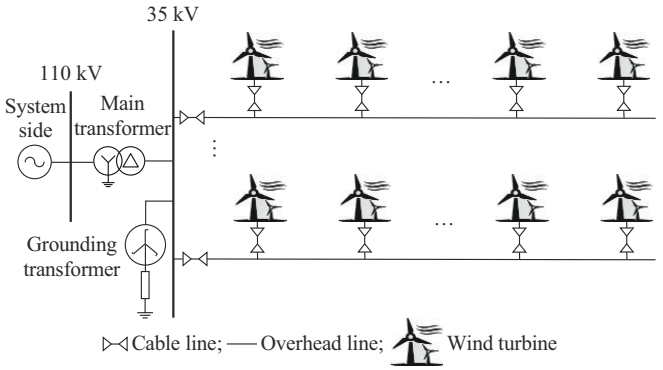

Fig. 1. Topological structure of electric system of a large-scale wind farm.

III. Simulation Model of DFIG-based Wind Farms and Generation of Training Sets

The PSCAD/EMTDC software is utilized to obtain single-phase grounding fault samples. The simulation model of a DFIG-based wind farm is shown in Fig. 2.

The simulation conditions are as follows. The sampling frequency is set to be 1600 Hz. There are 16 DFIGs with a single-unit capacity of 2 MW. In addition, the type of the cable lines is YJLV23-26/35kV-3 × 240 (3 cores with the cross-sectional area of 240 mm^2), and their lengths are all 100 m. The type of the overhead lines is LGJ240, and their lengths between adjacent wind turbines are shown in Fig. 2. The relevant parameters for the cable and overhead lines are listed in Table I.

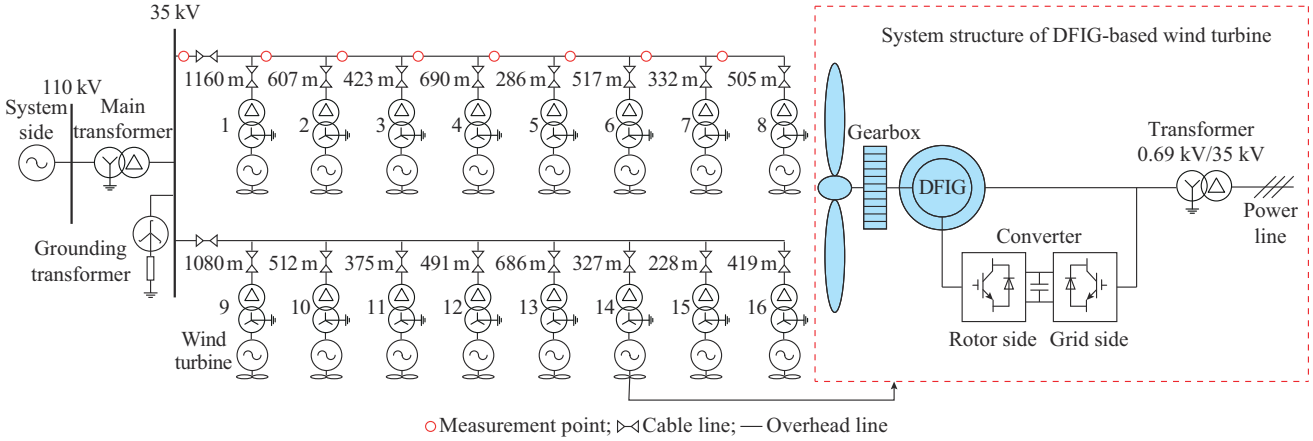

Fig. 2. Simulation model of a DFIG-based wind farm.

TABLE I
Related Parameters of Cable and Overhead Lines

Type	Sequence component	Resistance (Ω/km)	Reactance (Ω/km)	Susceptance (10^{-6} S/km)
YJLV23-26/ 35kV-3 × 240	Positive sequence	0.1250	0.1130	91.0000
	Zero sequence	1.2500	0.3955	57.0000
LGJ240	Positive sequence	0.1310	0.3720	4.0506
	Zero sequence	0.2810	1.1160	2.7946

Indeed, a training set of fault samples is crucial for generating a multi-RF fault locator. Therefore, appropriate and representative failure scenarios should be selected to obtain fault samples. The conditions for generating the training samples are as follows: ① assume that a fault occurs on the upper power line in Fig. 2; and ② consider the combinations of different fault locations and fault resistances when single-phase grounding short circuits occur. These generation conditions for the training set of fault samples are listed in Table II.

TABLE II
Generation Conditions for Training Set of Fault Samples

Fault parameter	Generation condition
Fault location (interval between the two adjacent single-phase grounding fault points in each section (m))	1 (230), 2 (120), 3 (80), 4 (130), 5 (50), 6 (100), 7 (60), 8 (100)
Fault resistance (Ω)	0, 25, 50, 75, 100

The fault resistances listed in Table II range from 0 Ω to 100 Ω, whereas the fault resistances related to the test set range from 0 Ω to 200 Ω.

IV. Methodology for Fault Characteristic Analysis and Fault Feature Extraction

A. Fault Characteristic Analysis

Signal processing is the key point and premise for feature extraction for fault location. As an efficient spectrum analysis method, the FFT algorithm [23] can rapidly transform the time domain into the frequency domain and accurately extract amplitude parameters. Thus, it is used in the characteristic analysis for fault signals of power lines of DFIG-based wind farms. If the complex exponential signal of a single frequency is selected as the input of the FFT algorithm, it can be expressed as:

$$x(n)=A_0\mathrm{e}^{\mathrm{i}(n\omega_0+\varphi_0)}=A_0\mathrm{e}^{\mathrm{i}(2\pi n\beta/N+\varphi_0)} \tag{1}$$

where N is the number of signal sampling points per cycle; β is the signal frequency index; $n\in[-N+1,N-1]$; A_0 is the amplitude; ω_0 is the angular frequency; φ_0 is the initial phase; and $2\pi\beta/N$ is the multiple forms of ω_0.

Through an FFT analysis, the calculated amplitude-frequency expression can be obtained as:

$$X(k)=\frac{A_0}{N}\frac{\sin(\pi(\beta-k))}{\sin(\pi(\beta-k)/N)}\mathrm{e}^{\mathrm{i}\left[\varphi_0+\frac{N-1}{N}(\beta-k)\pi\right]} \quad (2)$$

where k is an integer in the range of 0 to $N-1$.

It is obvious that the simulation model of the DFIG-based wind farm contains two power lines, each of which can be divided into eight sections. Additionally, distributed measurement points are installed at the head end of a section, as shown in Fig. 2. During the electromagnetic transient simulation, phase A grounding faults with transition resistances of 0, 25, 50, 75, and 100 Ω are set at different sections of the first power line of the DFIG-based wind farm. Taking the A-phase grounding fault that occurs at 1-230 m with a transition resistance of 100 Ω as an example, the waveform data at measurement point 1 and the FFT analysis results for the zero-sequence current are shown in Fig. 3.

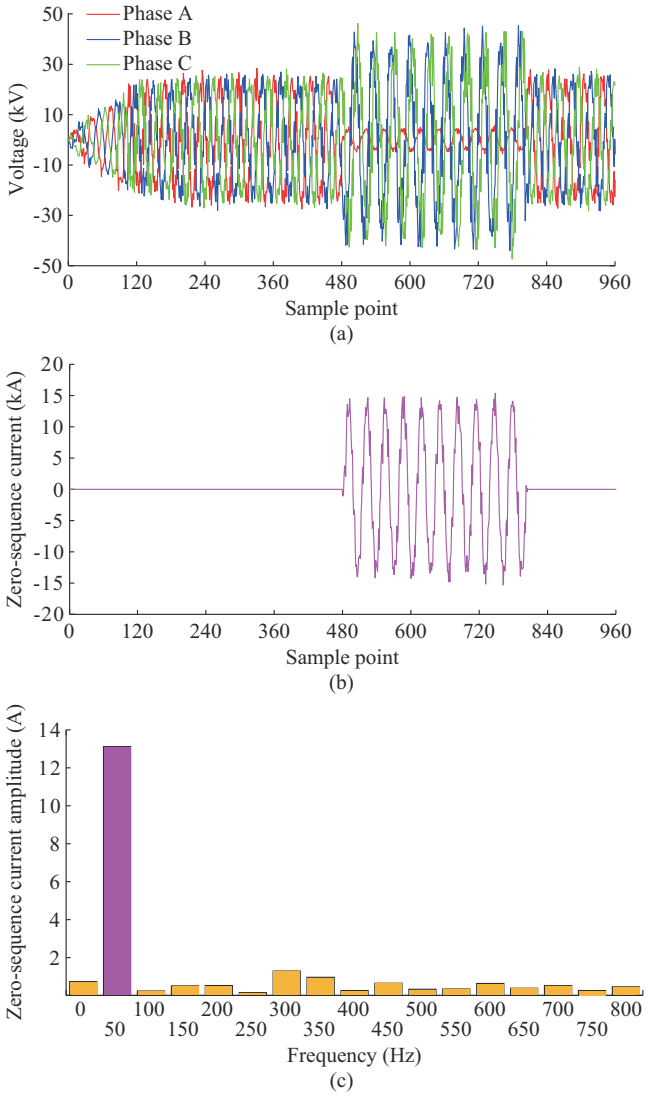

Fig. 3. Waveform data and FFT analysis results for zero-sequence current when a phase A grounding fault occurs at 1-230 m with a transition resistance of 100 Ω. (a) Three-phase voltage waveform. (b) Zero-sequence current waveform. (c) Each frequency component of zero-sequence current.

The notation "1-230" indicates that the fault point is at section 1 of the faulty line and is 230 m away from the 35 kV bus of the substation. More generally, *k-b* indicates that the fault point is at section k and the distance between the fault point and the 35 kV bus is b m. In fact, the measured data of the second cycle after the moment that a fault occurs are used to obtain accurate FFT results and avoid the unstable stage of the transient currents in the first cycle after a power line fault.

In particular, the purple bar in Fig. 3(c) shows that the amplitude of the fundamental frequency (50 Hz) of the zero-sequence current calculated by the FFT algorithm is 13.1224 A. Similarly, the values of the other measurement points can also be obtained in this manner. In conclusion, the fault characteristic analysis is effectively completed for the DFIG-based wind farm.

B. Fault Feature Extraction for DFIG-based Wind Farms

Data and features determine the upper bound of machine learning, whereas models and algorithms are only applied to approach this upper bound. Therefore, feature engineering is the key to establishing the location model. Moreover, the quality of the extracted features will have a direct impact on the fault location performance for DFIG-based wind farms.

It is necessary to extract fault features from the measured data of power lines of a DFIG-based wind farm because they are time series and large. From a practical point of view, wind turbines do not inject a zero-sequence current when a single-phase grounding fault occurs, because the windings at 35 kV side of all box-type transformers of the wind turbines are delta connections. As a result, their zero-sequence currents are not affected by the outputs of the wind turbines. In addition, an analysis of the feature importance based on RF algorithm shows that the features of the fundamental amplitudes of the zero-sequence currents can reflect the fault information and can be effectively used for fault location in power lines. In summary, the extraction of fault features is analyzed and designed according to Fig. 3. The FFT algorithm is used to extract the fundamental amplitudes of zero-sequence currents from waveform data measured at the measurement point when single-phase grounding faults occur, thereby constituting features 1 to 8 (F_1-F_8), as listed in Table III, where I_{k0} is the fundamental amplitude of the zero-sequence current at section k ($k=1, 2, \ldots, 8$).

TABLE III
FEATURES OF ZERO-SEQUENCE CURRENTS OF SINGLE-PHASE GROUNDING FAULTS FOR POWER LINES OF DFIG-BASED WIND FARMS

Feature label	Specific feature description	Feature label	Specific feature description
F_1	I_{10}	F_5	I_{50}
F_2	I_{20}	F_6	I_{60}
F_3	I_{30}	F_7	I_{70}
F_4	I_{40}	F_8	I_{80}

V. PROCEDURE TO CONSTRUCT MULTI-RF FAULT LOCATOR FOR DFIG-BASED WIND FARMS

The procedure to construct the multi-RF fault locator is as follows: ① model the problem of fault location as a classification and regression problem; ② sample historical fault da-

ta to train the optimal multi-RF fault locator; and ③ input the current sampled fault data into the optimal multi-RF fault locator to realize automatic location for power lines of DFIG-based wind farms.

A. RF Algorithm

The RF algorithm [19]-[22] constructs a supervised ensemble tree model whose base learners are all unpruned decision trees (DTs) constructed by the classification and regression tree (CART) algorithm. Moreover, it combines the ideas of the bagging and random subspace methods, which can effectively solve the bottleneck problems of single weak learner that are prone to overfitting and have difficulty in improving their performance.

Since a DT generated by the CART algorithm is used as the base learner to constitute the RF model, the classification strength s of a DT and the correlation ρ between DTs must be the main factors affecting the location performance, which can be characterized by the generalization error PE^*:

$$PE^* \le \frac{\bar{\rho}(1-s^2)}{s^2} \tag{3}$$

where $\bar{\rho}$ is the average value of ρ. When $\bar{\rho}$ is smaller and s is larger, the upper bound of PE^* will be lower. As a result, it strengthens the generalization ability and increases the location accuracy.

B. Feature Importance Theory Based on Node Impurity

During the training process for RF, the node impurity (Gini index) is used as the binary partition criterion to evaluate the segmentation effects when different features split the nodes. Suppose that node t of DT v is made up of subset $\boldsymbol{S}_t$ that includes a samples and n classes of C_k ($k=1,2,...,n$). In addition, a_k indicates the number of samples correctly classified in the subset after adopting a feature to split node t. Then, the original Gini index of $\boldsymbol{S}_t$ is given as:

$$Gini_v(\boldsymbol{S}_t) = 1 - \sum_{k=1}^{n} p_k^2 \tag{4}$$

where p_k is the probability that the sample points belong to class k and $p_k = a_k/a$. When an RF uses feature F_i to split node t, $\boldsymbol{S}_t$ is divided into subsets $\boldsymbol{S}_{t_1}$ and $\boldsymbol{S}_{t_2}$ according to the binary partition criterion. Thereafter, under the condition of feature F_i, the Gini index of $\boldsymbol{S}_t$ is expressed as:

$$Gini_v(\boldsymbol{S}_t, F_i) = \frac{a_{t_1}}{a} Gini_v(\boldsymbol{S}_{t_1}) + \frac{a_{t_2}}{a} Gini_v(\boldsymbol{S}_{t_2}) \tag{5}$$

where a_{t_1} and a_{t_2} are the numbers of samples in $\boldsymbol{S}_{t_1}$ and $\boldsymbol{S}_{t_2}$, respectively. From (4) and (5), the decrease in the node impurity before and after splitting node t is determined as:

$$\Delta Gini_v(\boldsymbol{S}_t, F_i) = Gini_v(\boldsymbol{S}_t) - Gini_v(\boldsymbol{S}_t, F_i) \tag{6}$$

On the basis of completing the construction of RF model, if feature F_i divides u nodes in DT v and carries out CART modeling for r DTs, then the mean decrease in the node impurity of the feature (feature importance) in the RF model is defined as:

$$\Delta_m Gini(F_i) = \frac{1}{r} \sum_{v=1}^{r} \sum_{t=1}^{u} \Delta Gini_v(\boldsymbol{S}_t, F_i) \tag{7}$$

Similarly, the mean decrease in the node impurities of all features can be computed using the above formulas. A larger mean decrease in the node impurity of a feature means that the role of that feature in the prediction models is greater. Therefore, this theory can be used to select features and quantitatively analyze the validity of the fault feature extraction to obtain a higher fault location accuracy.

C. Construction Process of Multi-RF Fault Locator

The corresponding steps for building the fault classification and regression models for power lines of DFIG-based wind farms using multi-RF algorithm are as follows.

Step 1: using the bagging method, w training sample subsets for fault classification are generated, which can be expressed as $\{\boldsymbol{S}_1, \boldsymbol{S}_2, ..., \boldsymbol{S}_w\}$. Then, the probability that a subset $\boldsymbol{S}_c$ ($c=1,2,...,w$) does not contain a certain sample point in Bootstrap sample can be obtained as:

$$p_{\text{OOB}} = \left(1 - \frac{1}{w}\right)^w \tag{8}$$

Its limit can be calculated as:

$$\lim_{w\to\infty} p_{\text{OOB}} = \lim_{w\to\infty} \left(1 - \frac{1}{w}\right)^w = \frac{1}{\text{e}} \approx 0.368 \tag{9}$$

Therefore, approximately 36.8% of the unsampled samples in $\boldsymbol{S}$ are called out-of-bag (OOB) data. Using the OOB data as a test set, the corresponding OOB error can be immediately evaluated once each CART base learner is constructed. Subsequently, an estimate of the generalization error of RF algorithm can be obtained by averaging the OOB errors of w DTs.

Step 2: for each training subset, a corresponding single DT is constructed according to the CART algorithm. The DTs built for w training sample subsets are denoted as $\{T_1, T_2, ..., T_w\}$ and used in combination to form an RF fault section locator.

Step 3: during the training of each DT, the random subspace method is used to split the nodes of the DT. Moreover, $m = \left\lfloor \sqrt{M} \right\rfloor$ sub-variables are randomly extracted from M variables with equal probability in the feature space to form a node-splitting candidate feature subset. For each candidate feature, its optimal splitting feature and splitting threshold are selected by (4) and (5) to divide a DT node until the DT stops growing according to the minimum principle of the Gini index.

Step 4: after each DT is constructed using the stop threshold method, the integrity of the tree is preserved without pruning. In addition, the predicted fault section sequence can be obtained for a test fault sample X by w DTs, which is $\{T_1(X), T_2(X), ..., T_w(X)\}$.

Step 5: the predicted fault section sequence obtained in *Step 4* is voted using (10), and the section with the highest number of votes on is selected as the final predicted fault section k of the RF fault section locator.

$$T(X) = \arg\max_{C_k} \sum_{j=1}^{w} I(T_j(X) = C_k) \quad k = 1,2,...,n \tag{10}$$

where $I(\cdot)$ is the indicator function. When the equality condition is true, the value of $I(\cdot)$ is 1; otherwise, it is 0. $T_j(X) =$

C_k means that the output classification result of DT j is C_k.

Meanwhile, during the training in *Steps 1-3*, the mathematical relationship between the OOB error and w DTs is defined as:

$$E_{\text{OOB}} = 1 - \frac{\sum_{a=1}^{N_O} I(T(O_a) = C_a)}{N_O} \tag{11}$$

where N_O is the total sample size of the OOB data; and $T(O_a)$ can be obtained in the form of (10), and $T(O_a) = C_a$ means that the OOB sample O_a is classified and the result is correct.

Step 6: with the above steps, an RF fault section locator can be built. Similarly, a multi-RF regression fault locator can be constructed to output the fault distance. The difference is that the average of the predicted values of all DTs is taken as the final location result for each RF regression fault locator.

VI. Design of Multi-RF Fault Locator for Power Lines of DFIG-based Wind Farms

The RF algorithm has the following advantages for solving the problem of predicting the fault distance of power lines of DFIG-based wind farms.

1) It possesses automatic parallel computing and a fast training speed and is suitable for processing large-scale power data.

2) Using the tree structure model, there is no need to normalize the fault data collected from DFIG-based wind farms.

3) Training sample subsets and features are randomly extracted using the bootstrap and random subspace methods, respectively, to minimize the correlation between DTs. Consequently, it has a strong generalization ability and high location accuracy.

The multi-RF algorithm has the advantages of the single-RF algorithm. Further, the multi-RF algorithm can improve the utilization of samples and achieve a higher location accuracy than the single-RF algorithm. In this paper, the multi-RF fault location framework for power lines of DFIG-based wind farms is designed according to Fig. 4. In this framework, a single-RF fault section locator based on the classification algorithm and a multi-RF regression fault locator based on the regression algorithm are constructed. The single-RF fault section locator is used to identify the sections where fault points are located. When the fault sections are determined, the switches in these sections are closed, whereas those in the other sections remain open. In the multi-RF regression fault locator, only the RFs corresponding to the fault sections are activated, and the predicted values of the fault distance can be output. In contrast, the RFs of other sections are in the disabled state, and no location results are output.

As shown in Fig. 4, the brief summary of the steps for the fault location of power lines of DFIG-based wind farms based on multi-RF algorithm is as follows.

Step 1: data preprocessing. A model of DFIG-based wind farms is established to obtain the zero-sequence current signals. Moreover, the fundamental amplitudes of the zero-sequence currents are extracted by the FFT algorithm in the MATLAB 2018a environment to construct the eight original features in Table III.

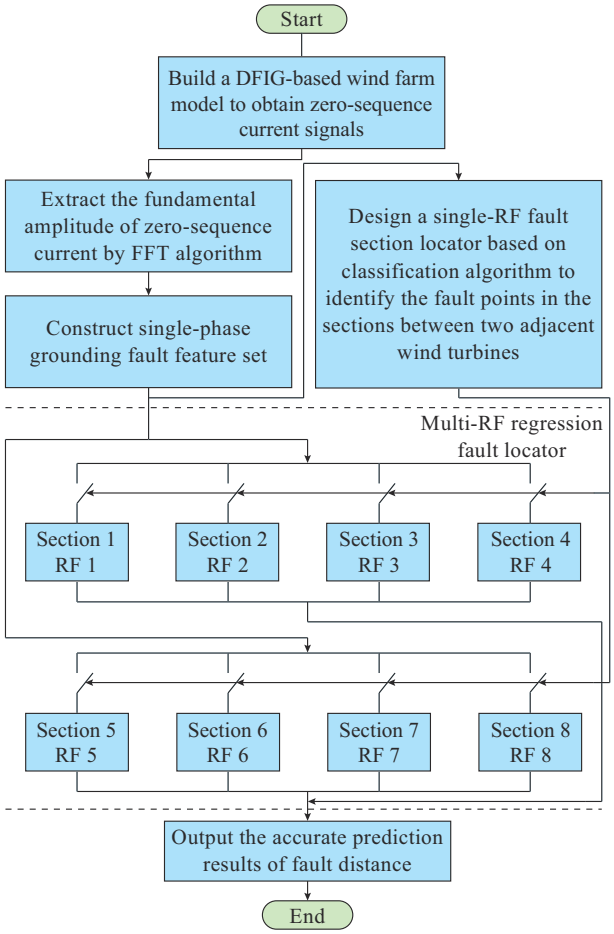

Fig. 4. Multi-RF fault location framework for power lines of DFIG-based wind farms.

Step 2: optimal classifier construction. The fault samples are input into a single-RF fault section locator for training, and test samples are used to identify the sections where the fault points are located. That is, the fault points in the sections between two adjacent wind turbines are locked.

Step 3: optimal multi-regressor construction. For each section, an RF regression fault locator is established. On this basis, *Step* 2 controls the on-off state of the switches in the flowchart to close the switches corresponding to the fault sections, thereby completing the prediction of the fault distance.

VII. Analysis and Discussion of Experimental Results

A. Evaluation Indicators

Criteria for evaluating the prediction methods are important for quantifying the location error. Therefore, an analysis of the model results is intended to evaluate the accuracy of the fault locator for the test set. Four performance indicators,

i.e., the prediction error (PE), the mean absolute error (MAE), the root-mean-square error (RMSE), and correlation coefficient (R) are selected as the evaluation indicators of the regression models. The mathematical definitions of these evaluation indicators are expressed as:

$$PE = y_{\text{pre},i} - y_{\text{act},i} \tag{12}$$

$$MAE = \frac{1}{N_t}\sum_{i=1}^{N_t}\left| y_{\text{pre},i} - y_{\text{act},i} \right| \tag{13}$$

$$RMSE = \sqrt{\frac{1}{N_t}\sum_{i=1}^{N_t}(y_{\text{pre},i} - y_{\text{act},i})^2} \tag{14}$$

$$R = corr(\boldsymbol{Y}_{\text{act}}, \boldsymbol{Y}_{\text{pre}}) = \frac{cov(\boldsymbol{Y}_{\text{act}}, \boldsymbol{Y}_{\text{pre}})}{\sigma_{\boldsymbol{Y}_{\text{act}}}\sigma_{\boldsymbol{Y}_{\text{pre}}}} \tag{15}$$

where $y_{\text{pre},i}$ and $y_{\text{act},i}$ are the predicted and actual values of the i^{th} fault location, respectively; N_t is the total number of test sample points; $corr(\cdot)$ is a correlation function; $\boldsymbol{Y}_{\text{act}}$ and $\boldsymbol{Y}_{\text{pre}}$ are the actual and predicted vectors of the fault location, respectively; $cov(\cdot)$ is a covariance function; and $\sigma_{\boldsymbol{Y}_{\text{act}}}$ and $\sigma_{\boldsymbol{Y}_{\text{pre}}}$ are the standard deviations of $\boldsymbol{Y}_{\text{act}}$ and $\boldsymbol{Y}_{\text{pre}}$, respectively.

The PE is introduced to measure the location error from the perspective of a single position, while the MAE, the RMSE, and R are used to evaluate the location and fitting effects from the overall perspective. In particular, R is the only criterion that should be close to 1 to reflect a better prediction accuracy.

B. Design of Optimal Multi-RF Fault Locator for DFIG-based Wind Farms

Before each RF algorithm is executed, the number of DTs n_{tree} needs to be initialized with a default value of 500. Then, the OOB error is utilized to evaluate the selection of n_{tree} to build the optimal multi-RF fault locator. The process for determining the number of DTs in an RF is shown in Fig. 5.

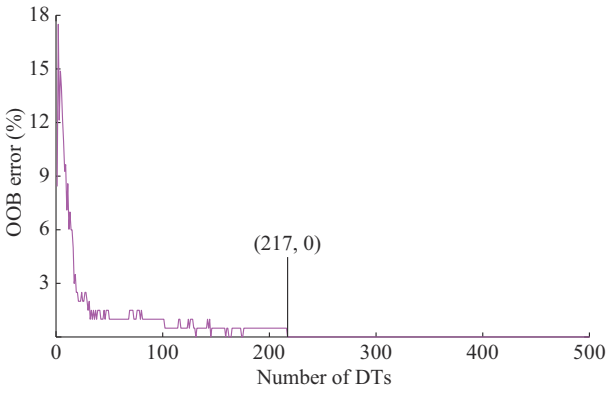

Fig. 5. Relationship between OOB error and number of DTs.

Figure 5 shows that the OOB error gradually decreases with the increase in the number of DTs when the number of DTs is less than 217. On the contrary, the OOB error tends to be stable; thus, the interval for the number of DTs is set as [217, 500]. Through testing and verification, taking the RF prediction accuracy and time cost as reference standards, and considering the complexity of RF structure, the optimal number of DTs is finally selected as 217 in the location stage.

The interpretability of the artificial intelligence model is an important index for the prediction of the fault distance, which can overcome the "black box" limitations of traditional artificial intelligence models. On the basis of determining the optimal number of trees, a DT in an RF is selected to display its splitting and growth processes. The optimal splitting features and thresholds of each node of a DT in an RF are listed in Table IV. Meanwhile, Fig. 6 shows the structure of a DT in an RF constructed by the divide-and-conquer stop threshold method.

TABLE IV
OPTIMAL SPLITTING FEATURES AND THRESHOLDS OF EACH NODE OF A DT IN AN RF

Node No.	Optimal splitting feature	Optimal splitting threshold	Node No.	Optimal splitting feature	Optimal splitting threshold
1	F_8	6.79275	12	F_1	16.55220
2	F_3	6.46630	13	—	—
3	—	—	14	—	—
4	F_1	15.53355	15	F_6	6.43825
5	F_7	6.31175	16	—	—
6	F_2	6.73475	17	—	—
7	F_7	0.08285	18	F_5	6.29805
8	F_4	6.46230	19	—	—
9	—	—	20	—	—
10	—	—	21	—	—
11	—	—			

Note: "—" indicates that a node is a leaf node.

C. Feature Importance Analysis Based on Optimal Multi-RF Algorithm

The feature importance score is directly related to the utilization rate of the features in the construction of the ensemble tree model, which can be characterized by the Gini index for feature splitting. According to Fig. 6, the mean decrease in the node impurity of a fault feature can be calculated to obtain the importance index of a single tree. Then, the calculated results of all trees are superimposed to obtain the feature importance score of the entire model, which is helpful for intuitive understanding and mining the relative importance of different quantities. In this paper, single-phase grounding fault points are set along different sections of the power line in the DFIG-based wind farm according to Table II, and five resistances of 0, 25, 50, 75, and 100 Ω are set at each fault point, which are used to train and generate the optimal multi-RF fault locator. Taking the construction of the RF fault section locator as an example, the feature importance score (mean decrease in node impurity) and sequence output from this process are shown in Fig. 7.

When a single-phase grounding fault occurs on the upper line in Fig. 2, the feature importance values of F_2-F_8 in Fig. 7 are relatively high. Therefore, F_2-F_8 can reflect the fault in-

formation and play an important role in the location of a fault section. Moreover, F_1 has a low feature importance value, which belongs to a redundant feature for the RF fault section locator mentioned in Section VI. However, it also plays a certain role in the training process of the model and can be used to identify faulty power lines. In this manner, the RF classification algorithm considers all of the above fault features to establish a fault section locator.

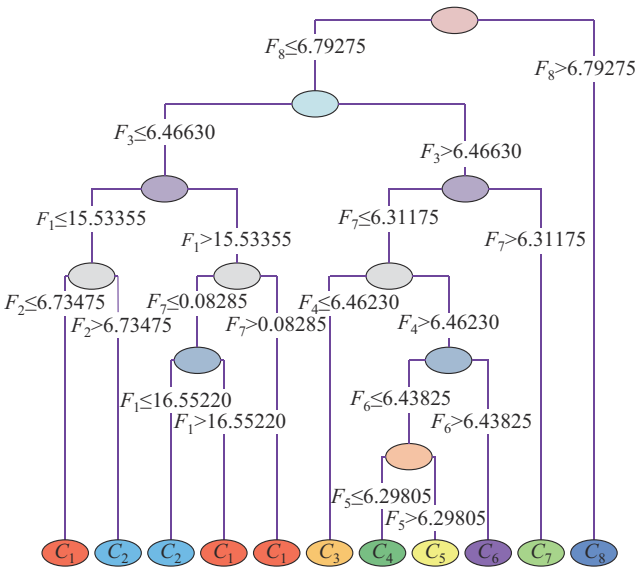

Fig. 6. Structure of a DT in an RF.

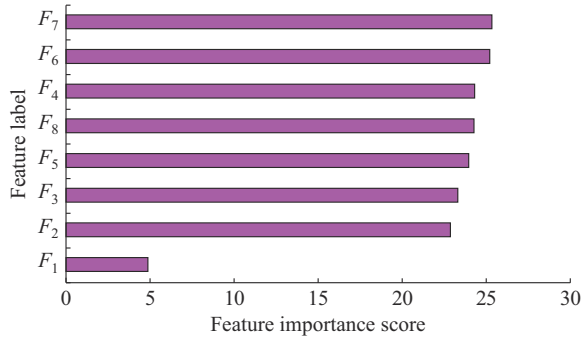

Fig. 7. Feature importance score and sequence.

D. Applicability Analysis for Power Lines of DFIG-based Wind Farms in Different Fault Scenarios

To verify the effect of location using the proposed method, the selected test samples should include all types of fault information and should be different from the training fault samples. Experiments are performed on a computer configured with an Intel(R) Core(TM) i7-4500U CPU running at 1.80 GHz, 8 GB RAM, and 64 bits operating system.

1) Comparison of Fault Section Location Using Different Algorithms

Considering that the fault data for an actual wind farm are not large, the total sample library of simulated fault cases consists of 230 sets to approach the actual situation. Of these 230 sets, 200 sets are used as training fault data (according to the generation conditions in Table II) to construct the fault section locator, and the remaining 30 sets are used as test fault data for recognition and location. The RF and support vector machine (SVM) algorithms, which use the same fault feature quantities extracted by the FFT algorithm, are compared and analyzed in terms of their accuracy and efficiency to illustrate the capability of locating a fault section for the RF algorithm. In the experiment for fault section location using the SVM, considering that the penalty factor C and radial basis function (RBF) parameter g directly affect the recognition accuracy, the grid search (GS) method is employed to optimize these two parameters with 10-fold cross-validation. The final optimized values of C and g are both 0.1. A comparison of the fault section location results using different algorithms is presented in Table V, where FN stands for fault number.

TABLE V
COMPARISON OF FAULT SECTION LOCATION RESULTS USING DIFFERENT ALGORITHMS

Classification algorithm	Test fault sample set	Location accuracy (%)	Operation time (s)
RF	T3 (FNs 1-30)	100	1.13
SVM	T3 (FNs 1-30)	100	12.33

From Table V, it can be concluded that both the RF and SVM algorithms can accurately identify the fault section for the power line in terms of the location accuracy, which demonstrates the effectiveness of feature extraction in this research. In terms of the location efficiency, the RF algorithm with a low computational complexity has higher classification efficiency than the SVM algorithm, reducing the operation time by 11.2 s in total. From the above comparative analysis, the RF algorithm is more suitable for locating a fault section in DFIG-based wind farms, which requires high real-time performance.

2) Analysis of Anti-fault Distance for Multi-RF Fault Locator

A phase A grounding fault with a transition resistance of 100 Ω is set on the power line of the DFIG-based wind farm, and the fault points are separately located at the same sections and different sections. Applying the single-RF fault section locator to classify the fault sections and trigger the corresponding switches to close, the prediction of fault points is carried out by the multi-RF regression fault locator. The experimental predicted results for different fault locations are listed in Table VI.

From Table VI, no matter where the fault occurs in the power line, the single-RF fault section locator can accurately lock the fault point in the corresponding section and successfully trigger the switch to close. The multi-RF regression fault locator ensures a high efficiency while making the predicted distance close to the true value. Furthermore, 1-60 m represents a fault that occurs on the cable line near the bus node, indicating that the proposed method can overcome the misjudgement of the cable line fault near the node. Therefore, the multi-RF fault locator can realize the automatic and accurate recognition of fault points, which is applicable to the faults on both cable and overhead line.

TABLE VI
EXPERIMENTAL PREDICTED RESULTS FOR DIFFERENT FAULT LOCATIONS

FN	Fault location (m)	Fault resistance (Ω)	Fault section classified by single-RF fault section locator	Multi-RF regression fault locator	
				Prediction result (m)	PE (m)
1	1-60	100	1	230.00	170.00
2	1-210	100	1	460.00	250.00
3	1-360	100	1	598.00	238.00
4	1-510	100	1	747.50	237.50
5	1-660	100	1	230.00	–430.00
6	1-810	100	1	575.00	–235.00
7	1-960	100	1	575.00	–385.00
8	1-1110	100	1	747.50	–362.50
9	1-200	100	1	230.00	30.00
10	2-1360	100	2	1424.73	64.73
11	3-1967	100	3	1940.83	–26.17
12	4-2390	100	4	2511.85	121.85
13	5-3080	100	5	2967.50	–112.50
14	6-3366	100	6	3401.08	35.08
15	7-3883	100	7	3904.00	21.00
16	8-4215	100	8	4381.67	166.67

TABLE VII
COMPARISON OF EFFECTS OF FAULT LOCATION WITH DIFFERENT FAULT LOCATORS AT DIFFERENT FAULT DISTANCES

FN	Single-RF regression		Single-SVR		Multi-SVR	
	Prediction result (m)	PE (m)	Prediction result (m)	PE (m)	Prediction result (m)	PE (m)
1	943.92	883.92	746.17	686.17	690.00	630.00
2	747.50	537.50	703.55	493.55	690.00	480.00
3	828.00	468.00	775.28	415.28	690.55	330.55
4	816.50	306.50	776.53	266.53	690.00	180.00
5	724.50	64.50	886.03	226.03	690.00	30.00
6	942.94	132.94	895.49	85.49	690.00	–120.00
7	1006.69	46.69	753.64	–206.36	690.00	–270.00
8	780.56	–329.44	1006.94	–103.06	690.00	–420.00
9	555.63	355.63	924.14	724.14	690.00	490.00
10	1398.25	38.25	1551.95	191.95	1523.94	163.94
11	1965.66	–1.34	2047.42	80.42	2006.98	39.98
12	2508.23	118.23	2534.72	144.72	2579.63	189.63
13	3146.70	66.70	2990.13	–89.87	3049.52	–30.48
14	3373.50	7.50	3492.37	126.37	3465.72	99.72
15	3968.50	85.50	3837.16	–45.84	3862.43	–20.57
16	4147.58	–67.42	4214.61	–0.39	4133.36	–81.64

3) Comparison of Multi-RF Regression Fault Locator with Single-RF Regression, Single-SVR, and Multi-SVR Fault Locators at Different Fault Distances

To further verify the superiority of the multi-RF regression fault locator for fault location, the single-RF regression, single support vector regression (SVR), and multi-SVR fault locators are used to analyze the effects of locating test samples with FNs 1-16 in the sample library, as listed in Table VII. A comparison of the PEs obtained by the four regression fault locators at different fault distances is shown in Fig. 8.

From Table VII and Fig. 8, it can be observed that for most faults, the location accuracy of the multi-RF regression fault locator is better than that of the single-RF regression, single-SVR, and multi-SVR fault locators, and the PE can be decreased by up to 713.92 m. Moreover, the effects of location are not significantly affected by different fault distances, and the stability is high, which can realize the reliable location of fault points in complex DFIG-based wind farms.

4) Analysis of Anti-fault Resistance for Multi-RF Fault Locator

Considering that the actual faults of power lines in DFIG-based wind farms are usually nonmetallic short circuits, there is often a transition resistance at each fault point. The resistance is usually uncertain and has a diversified trend, i.e., it significantly varies in different fault situations. Therefore, it is necessary to verify whether the proposed method is applicable for testing fault samples with different fault resistances. The experimental results for multi-RF fault locator when phase A grounding faults occur with different fault resistances are listed in Table VIII, where ① indicates that the fault resistance is the same as the training sample set, however, the fault location is the opposite; and ② indicates that the fault resistance and fault location are different from the training sample set.

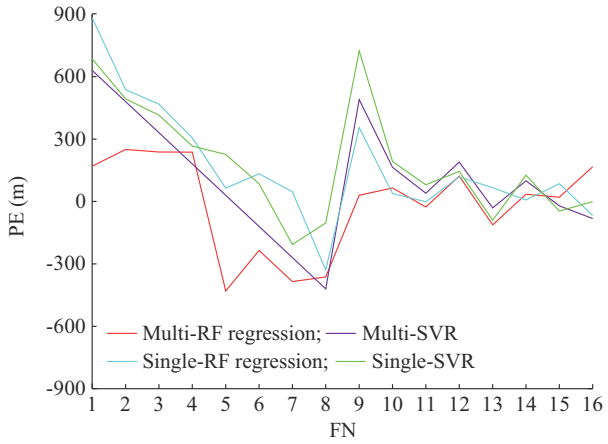

Fig. 8. Comparison of PEs obtained by four regression fault locators at different fault distances.

From Table VIII, in particular, with respect to ②, the single-RF fault section locator can still correctly identify the sections corresponding to nonmetallic faults without increasing the original training data. Switches are closed to start the multi-RF regression fault locator. In addition, the prediction results obtained by the optimal multi-RF regression fault locator can maintain the accuracy at higher fault resistance values. For example, when the fault resistance is 200 Ω and the fault point is located at 8-4400 m, the identification result is C_8, and the location section is 8, which coincides with the real fault section. On this basis, the obtained prediction error of the fault distance is only –18.33 m. The results show that the multi-RF fault locator is insensitive to the changes in the fault resistance and can be applied to an actual wind farm system.

TABLE VIII
EXPERIMENTAL RESULTS FOR MULTI-RF FAULT LOCATOR WHEN PHASE A GROUNDING FAULTS OCCUR WITH DIFFERENT TRANSITION RESISTANCES

Case of testing fault samples	FN	Fault resistance (Ω)	Fault location (m)	Fault section classified by single-RF fault section locator	Multi-RF regression fault locator	
					Prediction result (m)	PE (m)
	17	0	1-400	1	230.00	−170.00
	18	25	1-400	1	345.00	−55.00
①	19	50	2-1566	2	1647.27	81.27
	20	75	2-1566	2	1518.91	−47.09
	21	100	3-2020	3	2081.00	61.00
	22	20	4-2708	4	2523.27	−184.73
	23	40	4-2708	4	2487.92	−220.08
	24	60	5-2935	5	2930.00	−5.00
	25	80	5-2935	5	2967.50	32.50
②	26	120	6-3413	6	3413.15	0.15
	27	140	6-3413	6	3459.49	46.49
	28	160	7-3820	7	3808.82	−11.18
	29	180	7-3820	7	3812.67	−7.33
	30	200	8-4400	8	4381.67	−18.33

5) Comparison of Multi-RF Regression Fault Locator with Single-RF Regression, Single-SVR, and Multi-SVR Fault Locators for Different Fault Resistances

Table IX compares the effects of fault location for the test samples with FNs 17-30 in the sample library on the premise of extracting the same fault feature quantities by different fault locators. Furthermore, to further illustrate the ability of multi-RF regression fault locator to resist transition resistances when locating faults, Fig. 9 shows a comparison of the PEs of multi-RF regression fault locator with single-RF regression, single-SVR, and multi-SVR fault locators for different fault resistances.

TABLE IX
COMPARISON OF EFFECTS OF FAULT LOCATION WITH DIFFERENT FAULT LOCATORS FOR DIFFERENT FAULT RESISTANCES

FN	Single-RF regression		Single-SVR		Multi-SVR	
	Prediction result (m)	PE (m)	Prediction result (m)	PE (m)	Prediction result (m)	PE (m)
17	449.94	49.94	884.07	484.07	689.96	289.96
18	969.31	569.31	664.48	264.48	690.72	290.72
19	1698.35	132.35	1502.97	−63.03	1538.40	−27.60
20	1655.08	89.08	1527.23	−38.77	1532.05	−33.95
21	1997.13	−22.87	2027.85	7.85	2006.98	−13.02
22	2623.96	−84.04	2581.88	−126.12	2579.62	−128.38
23	2522.04	−185.96	2579.68	−128.32	2579.63	−128.37
24	3070.10	135.10	3027.01	92.01	3001.65	66.65
25	3185.17	250.17	2980.06	45.06	3027.23	92.23
26	3494.66	81.66	3403.90	−9.10	3465.72	52.72
27	3529.66	116.66	3283.19	−129.81	3465.72	52.72
28	3886.46	66.46	3639.67	−180.33	3862.43	42.43
29	3871.96	51.96	3693.30	−126.70	3862.43	42.43
30	4017.08	−382.92	3505.86	−894.14	4311.57	−88.43

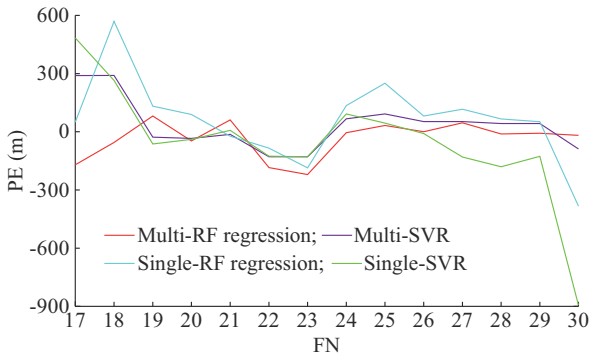

Fig. 9. Comparison of PEs obtained by four regression fault locators for different fault resistances.

From Table IX and Fig. 9, it can be observed that among the four regression fault locators, the multi-RF regression fault locator has the highest level of anti-fault resistance. For example, the location result is 8 when the fault resistance is 200 Ω and the fault point is located at 8-4400 m. Compared with the PEs of single-RF regression, single-SVR, and multi-SVR fault locators, the PEs are reduced by 364.59, 875.81, and 70.10 m, respectively. In conclusion, the multi-RF regression fault locator can better adapt to the trend of uncertainty and diversification of fault resistances and ensure the accuracy in locating the faults in DFIG-based wind farms.

The average performance of data sets T1 (FNs 1-16), T2 (FNs 17-30), and T3 (FNs 1-30) with different numbers of samples is tested using four fault locators, respectively. The final results are shown in Fig. 10.

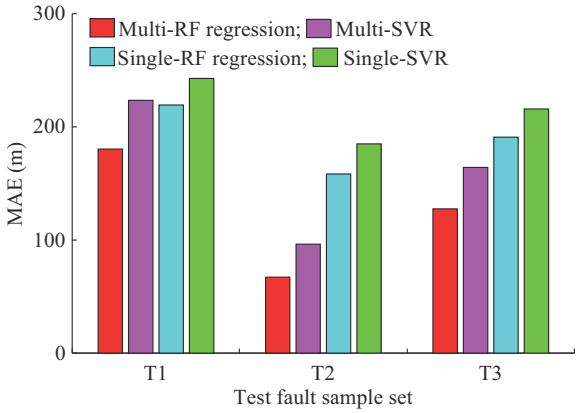

Fig. 10. Comparison of MAEs obtained by four regression fault locators using different test sets.

It can be observed from Fig. 10 that the MAEs of the multi-RF regression and multi-SVR fault locators are lower than those of the single-RF regression and single-SVR fault locators when locating faults. Moreover, the multi-RF regression fault locator is better than the multi-SVR fault locator, and the MAEs of T1, T2, and T3 are reduced by 43.15, 29.25, and 36.66 m, respectively. These results demonstrate the validity and applicability of the multi-RF regression fault locator in different fault scenarios for DFIG-based wind farms.

The fitting performance of data sets T1, T2, and T3 is test-

ed using the four fault locators, respectively. The calculated RMSEs and *R* are presented in Tables X and XI. From these tables, the results for the multi-RF regression fault locator are the best. Thus, the proposed multi-RF regression fault locator is suitable for the regression analysis of power line data of DFIG-based wind farms to obtain accurate fault points.

TABLE X
CALCULATED RMSES OF FOUR FAULT LOCATORS FOR DIFFERENT DATA SETS

Data set	Calculated RMSE (m)			
	Multi-RF regression	Multi-SVR	Single-RF regression	Single-SVR
T1	221.77	291.12	323.46	324.45
T2	96.65	129.16	215.32	294.92
T3	174.90	230.19	278.27	311.02

TABLE XI
CALCULATED *R* OF FOUR FAULT LOCATORS FOR DIFFERENT DATA SETS

Data set	Calculated *R*			
	Multi-RF regression	Multi-SVR	Single-RF regression	Single-SVR
T1	0.9871	0.9808	0.9821	0.9848
T2	0.9974	0.9958	0.9873	0.9815
T3	0.9923	0.9886	0.9863	0.9840

6) Analysis of Anti-noise Properties of Multi-RF Fault Locator

The measured data of a DFIG-based wind farm are in time series, usually accompanied by noise or electromagnetic interference. The fault location model should have good robustness in different noise environments and be able to accurately predict the fault distance. For this purpose, Table XII lists the location results at 8-4400 m with a fault resistance of 200 Ω and signal-to-noise ratios (SNRs) of 50, 40, and 30 dB. It can be observed from Table XII that the multi-RF fault locator is capable of dealing with noisy data and is robust against noise and electromagnetic interference. Since the fundamental frequency components of the zero-sequence currents in the second cycle after the faults are used, the noise has no effect on the PE of the fault distance.

TABLE XII
LOCATION RESULTS AT 8-4400 M WITH DIFFERENT SNRS USING MULTI-RF FAULT LOCATOR

SNR (dB)	Fault section classified by single-RF fault section locator	Multi-RF regression fault locator	
		Prediction result (m)	PE (m)
Without noise	8	4381.67	–18.33
50	8	4381.67	–18.33
40	8	4381.67	–18.33
30	8	4381.67	–18.33

7) Comparison of Multi-RF Method with Other Methods

An experiment is carried out using the multi-RF method and data set T3, which contains fault attributes such as the fault location and fault resistance, as listed in Tables VI and VIII. The proposed method is compared with three previously reported methods. These earlier methods are based on the dual-time transform and DT (method 1) [24], redundancy parameter estimation (method 2) [1], and the dual-ended traveling wave (method 3) [1]. A comparison of the proposed method with these methods is presented in Table XIII. In addition, a 1 μs synchronization error results in a location error of at least 300 m for the synchronization problem. Therefore, the latter two methods must consider the impact of synchronization. It is necessary to add 300 m to the original location error to suit the actual situation. The results in Table XIII clearly show that the overall efficacy of the proposed method is superior to that of previously reported methods for locating fault sections and predicting fault distances.

TABLE XIII
COMPARISON OF PROPOSED METHOD WITH OTHER REPORTED METHODS

Method	Sampling frequency for fault section location (Hz)	Sampling frequency for fault location (Hz)	The maximum fault resistance (Ω)	Location accuracy (%)	Does synchronization need to be considered?	MAE (m)
1	1000	10000	100	99.96	No	2624.00
2	1000000	1000000	None	None	Yes	362.92
3	1000000	1000000	None	None	Yes	467.83
Proposed	1600	1600	200	100.00	No	127.54

VIII. CONCLUSION

With the increasing use of wind farms, an intelligent single-phase grounding fault location method for power lines based on multi-RF algorithm is proposed in this paper. The main innovations and conclusions of this paper are as follows.

1) An FFT algorithm is introduced to extract the fundamental amplitudes of the zero-sequence currents to rapidly construct the original feature set, which is not affected by DFIGs and has a high reliability.

2) A quantitative analysis of the feature importance based on node impurity is designed, which verifies the effectiveness of feature extraction.

3) PSCAD/EMTDC software is used to produce the samples of different and enriched fault scenarios, and a multi-RF fault locator is constructed for fault location. Compared with the MAEs of single-RF regression, single-SVR, and multi-SVR fault locators, the MAEs are reduced by 63.41, 88.33, and 36.66 m, respectively, on data set T3.

4) A comparison between the proposed and previously reported methods shows that the proposed method is more feasible for locating the multi-branch and hybrid short lines of a DFIG-based wind farm.

Although the proposed method has distinct advantages in locating the single-phase grounding fault for power lines of DFIG-based wind farms, it has certain limitations. For instance, the fundamental amplitudes of the zero-sequence currents are used as the feature quantities in this method, which makes it impossible to locate three-phase symmetrical faults. Therefore, future work will focus on improving the method to further enhance the applicability of fault location.

REFERENCES

[1] K. Zhang, Y. Zhu, and X. Liu, "A fault locating method for multi-branch hybrid transmission lines in wind farm based on redundancy parameter estimation," *Journal of Modern Power Systems and Clean Energy*, vol. 7, no. 5, pp. 1033-1043, Sept. 2019.

[2] J.-W. Lee, W.-K. Kim, J. Han *et al*., "Fault area estimation using traveling wave for wide area protection," *Journal of Modern Power Systems and Clean Energy*, vol. 4, no. 3, pp. 478-486, Jul. 2016.

[3] J. Zhao, H. Hou, Y. Gao *et al*., "Single-phase ground fault location method for distribution network based on traveling wave time-frequency characteristics," *Electric Power Systems Research*, vol. 186, pp. 1-9, Sept. 2020.

[4] M. Korkali, H. Lev-Ari, and A. Abur, "Traveling-wave-based fault-location technique for transmission grids via wide-area synchronized voltage measurements," *IEEE Transactions on Power Systems*, vol. 27, no. 2, pp. 1003-1011, May 2012.

[5] M. A. Aftab, S. M. S. Hussain, I. Ali *et al*., "Dynamic protection of power systems with high penetration of renewables: a review of the traveling wave based fault location techniques," *International Journal of Electrical Power & Energy Systems*, vol. 114, pp. 1-13, Jan. 2020.

[6] X. Liu, D. Wang, X. Jiang *et al*., "Fault location algorithm for distribution power network based on relationship in time difference of arrival of traveling wave," *Proceedings of the CSEE*, vol. 37, no. 14, pp. 4109-4115, Jul. 2017.

[7] R. J. Hamidi and H. Livani, "Traveling-wave-based fault-location algorithm for hybrid multiterminal circuits," *IEEE Transactions on Power Delivery*, vol. 32, no. 1, pp. 135-144, Feb. 2017.

[8] S. Zhang, S. Lin, Z. He *et al*., "Ground fault location in radial distribution networks involving distributed voltage measurement," *IET Generation*, *Transmission & Distribution*, vol. 12, no. 4, pp. 987-996, Feb. 2018.

[9] K. Jia, T. Bi, Z. Ren *et al*., "High frequency impedance based fault location in distribution system with DGs," *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 807-816, Mar. 2018.

[10] J. Weng, D. Liu, N. Luo *et al*., "Distributed processing based fault location, isolation, and service restoration method for active distribution network," *Journal of Modern Power Systems and Clean Energy*, vol. 3, no. 4, pp. 494-503, Dec. 2015.

[11] Z. Li, Y. Ye, X. Ma *et al*., "Single-phase-to-ground fault section location in flexible resonant grounding distribution networks using soft open points," *International Journal of Electrical Power & Energy Systems*, vol. 122, pp. 1-10, Nov. 2020.

[12] M. A. Elsadd and A. Y. Abdelaziz, "Unsynchronized fault-location technique for two- and three-terminal transmission lines," *Electric Power Systems Research*, vol. 158, pp. 228-239, May 2018.

[13] J. Chen, E. Chu, Y. Li *et al*., "Faulty feeder identification and fault area localization in resonant grounding system based on wavelet packet and Bayesian classifier," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 4, pp. 760-767, Jul. 2020.

[14] V. L. Merlin, R. C. D. Santos, S. L. Blond *et al*., "Efficient and robust ANN-based method for an improved protection of VSC-HVDC systems," *IET Renewable Power Generation*, vol. 12, no. 13, pp. 1555-1562, Oct. 2018.

[15] H. Livani and C. Y. Evrenosoglu, "A machine learning and wavelet-based fault location method for hybrid transmission lines," *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 51-59, Jan. 2014.

[16] Z. Moravej, M. Movahhedneya, and M. Pazoki, "Gabor transform-based fault location method for multi-terminal transmission lines," *Measurement*, vol. 125, pp. 667-679, Sept. 2018.

[17] I. M. Karmacharya and R. Gokaraju, "Fault location in ungrounded photovoltaic system using wavelets and ANN," *IEEE Transactions on Power Delivery*, vol. 33, no. 2, pp. 549-559, Apr. 2018.

[18] C. Li, Z. Liu, Y. Zhang *et al*., "Diagnosis and location of the open-circuit fault in modular multilevel converters: an improved machine learning method," *Neurocomputing*, vol. 331, pp. 58-66, Feb. 2019.

[19] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5-32, Oct. 2001.

[20] M. V. Reddy and R. Sodhi, "A modified S-transform and random forests-based power quality assessment framework," *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 1, pp. 78-89, Jan. 2018.

[21] S. Bag, A. K. Pradhan, S. Das *et al*., "S-transform aided random forest based PD location detection employing signature of optical sensor," *IEEE Transactions on Power Delivery*, vol. 34, no. 4, pp. 1261-1268, Aug. 2019.

[22] C. Li, Y. Tao, W. Ao *et al*., "Improving forecasting accuracy of daily enterprise electricity consumption using a random forest based on ensemble empirical mode decomposition," *Energy*, vol. 165, pp. 1220-1227, Dec. 2018.

[23] J. Li, Q. Yang, H. Mu *et al*., "A new fault detection and fault location method for multi-terminal high voltage direct current of offshore wind farm," *Applied Energy*, vol. 220, pp. 13-20, Jun. 2018.

[24] S. Biswas, P. K. Nayak, and G. Pradhan, "A dual-time transform assisted intelligent relaying scheme for the STATCOM-compensated transmission line connecting wind farm," *IEEE Systems Journal*. doi: 10.1109/JSYST.2021.3070448

Yongli Zhu received the B.S., M.S., and Ph.D. degrees all in electrical engineering from North China Electric Power University (NCEPU), Beijing, China, in 1984, 1987, and 1992, respectively. He is currently a Professor and a Ph. D. Supervisor at NCEPU. His research interests include power equipment condition monitoring and analysis, power system analysis and control, power system network monitoring, and intelligent information processing.

Hua Peng received the B.S. degree in electrical engineering from Shandong Agricultural University, Tai'an, China, in 2015, and the M.S. degree in electrical engineering from Northeast Electric Power University, Jilin, China, in 2018. He is currently pursuing the Ph.D. degree in electrical engineering at North China Electric Power University, Beijing, China. His research interests include fault analysis and fault location of power lines in wind farm.